

健康相關生活品質之效用測量方法信度與效度的評估：以血液透析之末期腎病患者為例

林榮第¹ 姚開屏²

游芝亭³ 王榮德³

健康相關生活品質的測量已逐漸被運用於臨床決策、決定衛生政策優先順序、評估衛生計畫效益及經濟學之成本—效用分析上。本研究首先嘗試以血液透析患者為研究對象，對時間交換及評價等級兩種效用測量方法在健康相關生活品質之測量上的信度及效度進行評估，希望能建立起一個適合我國國情的健康狀態效用評估方法，並逐步修改而推廣到其他健康狀態的效用評估上，以作為臨床決策、衛生政策分析及計畫評估的參考指標。本研究採橫斷研究法，以1993年8月至11月期間，在臺北市三所教學醫院之血液透析中心接受長期血液透析，能夠並願意接受訪問的所有末期腎病患者為研究對象。由標準化訪視員以結構式問卷進行訪視及再測。共有125名患者完成訪問，回應率為90%。研究結果發現，時間交換及評價等級兩種效用測量方法對於不同的訪視者、不同的訪視時間及不同的訪視方法或訪視情境，都可得到穩定的結果，而且兩種方法在建構效度及效標相關效度等效度檢定上，均具有良好的表現。標準賭博、時間交換及評價等級等方法所測得的血液透析患者之健康相關生活品質的效用數值分別為0.75、0.61與0.57。

在健康狀態效用的測量方法中，標準賭博、時間交換及評價等級皆具有良好的信度及效度。但就可行性而言，時間交換及評價等級是較受大多數學者所建議採行的方法。另外，由於評價等級施測容易且所得結果與其他方法同樣具有信度及效度，故很適合應用於大規模的研究。（中華衛誌 1997；16(5)：404-416）。

關鍵詞：健康相關生活品質、效用、信度、效度

前言

隨著人們在健康方面需求的增加，投資在醫療衛生上的費用正逐年增加中。以美國而言，醫療費用所佔國民總生產毛額的比例，已由1960年的5.2%增加到1990年的12.2% (1)。醫療費用的年上漲率遠超出國民總生產毛額的成長率。在醫療費用高漲、醫療資源有限的情況下，如何對各項衛生政策之重要性與效益進行正確的評估，使其能發揮最大的功能，是政府決策者所面臨的一大難題。

¹臺北市立和平醫院內科部

²台大心理系

³台大公共衛生學院職業醫學與工業衛生研究所

聯絡人：王榮德

聯絡地址：臺北市仁愛路一段一號1443室，臺大公衛學院職業醫學與工業衛生研究所

聯絡電話：(02)3562224

傳 真：(02)3224660

投稿日期：85年6月13日

接受日期：86年8月6日

近年來，健康狀態及生活品質的測量逐漸被廣泛運用於臨床醫療、臨床及流行病學研究、衛生政策分析及計畫評估與人群監視等方面的研究上[2]。生活品質乃是人類經由日常生活的活動中所獲得的一種深度的滿足感，其高度涉及個人的主觀價值判斷[3]。換言之，生活品質是個人或社會對其生活經驗所給予的價值判斷，它包含所有與生活有關之人類的經驗、狀態、知覺與想法。因此，生活品質包括文化、生理、精神、人際關係、經濟、政治、宗教及哲學等層面[4]。因為生活品質所涵蓋的層面相當廣泛，故建議在使用此名詞時，最好指出所欲探討的層面。而與衛生政策決策分析最相關的生活品質為健康相關生活品質(HRQL, health-related quality of life)。在此層面之生活品質的定義大致可歸納為生活需要的滿意度[5]，與健康相關的感受經驗[6]或心理、社會及生理方面的健康狀態[7]。健康相關生活品質是指生活品質中與健康有關的部份[8]。健康相關生活品質的測量方式大致可分為整體性觀點(global approach)與多向度觀點(multidimensional approach)兩種方式[9]。根據世界衛生組織的定義[10]，健康包括生理性、心理性及社會性三個向度。因此，由多向度觀點來測量健康相關生活品質時，可由生理性、心理性及社會性三個向度作考量，依據各向度所應包含的主要評估項目[11-13]，選取具有信度及效度的臨床量表，來評估病患在三個不同向度的健康相關生活品質。綜合此三向度的評估結果，我們便可由多向度觀點來測量個體的健康相關生活品質。

整體性測量方法則是以一個健康狀態的綜合數值來代表個體的健康相關生活品質。在整體的健康相關生活品質的測量方法中，效用測量方法(utility approach)(8)是一種簡單可行且具有效度及信度的方法。效用方法所測得的結果是一個介於0與1之間的數值，其代表著某特定時間點上個體的健康相關生活品質。健康狀態效用的測量方法包括標準賭博法(standard gamble, SG)、時間交換法(time trade-off, TTO)、評價等級法(rating scale, RS)三種方法。在西方文化上，多數學者認為效

用測量方法在健康相關生活品質的測量上具有很好的信度及效度[14]。至於在華人方面，並沒有學者曾對健康狀態效用測量方法的信度及效度進行檢定。本研究首先嘗試以血液透析患者為研究對象，對TTO及RS兩種效用測量方法的信度及效度進行評估，並將效用的觀念運用於健康相關生活品質的測量上，希望能建立起一個適合我國文化背景的健康效用評估方法，並逐步修改而推廣到其他健康狀態的效用評估上，以期作為臨床決策、衛生政策分析及計畫評估的參考指標。

材料與方法

本研究係採橫斷研究法(cross-sectional study)，探討如何由效用的觀點來測量健康相關生活品質，並對此測量方法的信度與效度進行評估。茲分別就研究對象、研究工具、標準化訪視工作及健康相關生活品質效用測量方法的信度及效度的評估逐一說明如下：

研究對象

為避免施測對象的病情對施測方法之信度及效度的評估造成影響，本研究首先嘗試以意識狀態較不受疾病影響且病程較穩定之血液透析患者為施測對象。所收集的研究對象為1993年8月至11月期間，在臺北市三所教學醫院的血液透析中心接受長期血液透析(maintenance hemodialysis)，能夠並願意接受訪問的所有末期腎臟疾病患者。對於連續二次均拒絕訪問或因為意識障礙(Glasgow昏迷指數在15分以下)[15]、失明、身體狀況甚差(ECOG等級為4級)[16,17]而無法接受訪問的患者，則不符合本研究之收案條件。

研究工具

本研究中與健康有關之生活品質的測量資料，係根據結構式問卷訪視調查而來。問卷在經過前測(pretest)修改和復試後，才正式使用於訪視調查。問卷的內容共分三部份，茲概述如下：

- (一) 基本資料：此部份包括可能與健康相關生活品質有關的人口統計學資料及臨床醫學資料(18-21)。

(二) 整體性健康相關生活品質的效用測量方法：此部份包括三種由效用觀點來測量健康相關生活品質的方法[8,14]，即SG、TTO及RS。

SG[22,23](圖一)：此方法令受試者在一個賭局及一個確定結果所組成的兩個選擇項間做選擇。借助機率轉盤(23)來幫助受試者以機率的思索方式作選擇。藉著調整選擇項的機率大小來測量受試者對該健康狀態喜好程度的效用值。當受試者在賭局及確定結果間無法做選擇時，可計算出受試者對此健康狀態喜好程度的效用值。

TTO[24,25](圖二)：此方法令受試者在兩個具有確定結果的選項間作選擇。其中一個選項為維持目前的健康狀態直到死亡，另一個選項為完全健康的活一段較短的時間。藉著調整所願用以交換疾病狀態的完全健康日數來測量受試者對該健康狀態喜好程度的效用值，並借助類似圖二之視覺輔助工具(TTO with visual aid, TTOVA)[25]以幫助受試者了解測試內容。當受試者在兩個選項間無法作選擇時，由所願用以交換的完全健康日數即可計算得到該健康狀態的效用值。

RS[26](圖三)：此方法令受試者依其對健康狀態的喜好程度在一條兩端標示著0(死亡)與1(完全健康)的線上作標示。由標示點到原點的距離來計算受試者對該健康狀態喜好程度的效用值。

(三) 多向度健康相關生活品質量表：即以生

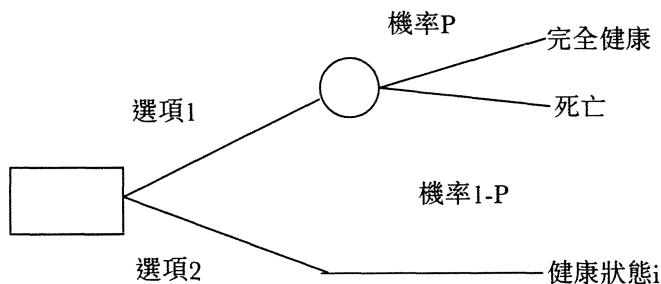
理性、心理性及社會性三個向度做考量，兼採一般性與專一性量表的特點，選取臨床上廣泛使用且具有信度及效度的量表，並根據前測結果修改「疼痛狀態量表」、「身體症狀」與「疲勞症狀」部份內容，以便更精確地量化，由多向度觀點來評估病患的健康相關生活品質。量表中各向度所包含的主要評估項目的選取主要依據下列原則[11-13]：生理向度主要包含與疾病或治療有關的症狀、身體活動能力與自我照顧能力等項目；社會向度主要包含社會角色功能、社會支持功能與人際間接觸功能等項目；心理向度主要包含精神狀態與認知功能等項目。

生理向度生活品質量表：包括修改式卡式評分等級(modified Karnofsky scale, ECOG scale)[16,17]、心臟功能分類[27]、呼吸功能量表[28]、疲勞狀態量表[29]、身體症狀量表[30]、疼痛狀態量表[31]以及Barthel氏指數(Barthel's index) [32]。

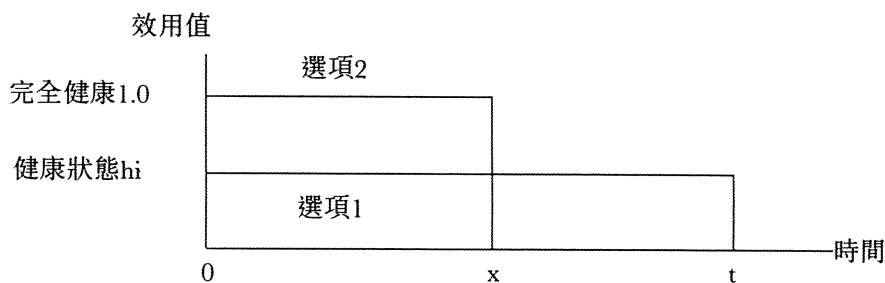
社會向度生活品質量表：包括社會角色功能量表[33]、社會支持功能量表[34]、以及人際間接觸量表[34]。

心理向度生活品質量表：包括精神健康狀態量表[36]、意識狀態量表[15]以及迷你精神狀態檢查量表[37]。

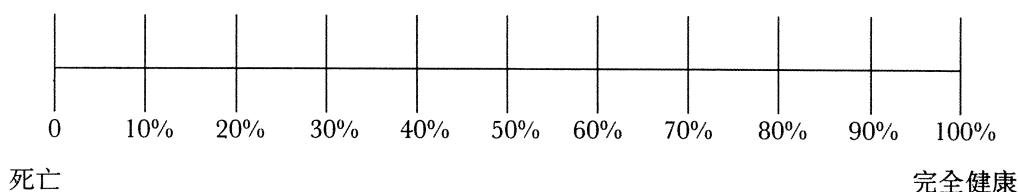
上述各向度所包含的生活品質量表之給分與其所代表的健康相關生活品質間關係如下：Barthel氏指數、意識狀態量表及迷你精神狀態檢查量表的得分與其所代表的健康相關



圖一 比死亡好之慢性健康狀態的標準賭博方法



圖二 比死亡好之慢性健康狀態的時間交換方法
 h_i : 健康狀態 i , t : 生病存活時間 , x : 健康存活的交換時間



圖三 比死亡好之慢性健康狀態的評價等級方法

生活品質間呈正相關性；其餘量表的得分與其所代表的健康相關生活品質間則呈負相關性。

標準化訪視工作

訪視工作係由二位研究生擔任。訪視前並進行研究計劃之說明、問卷內容之講解、訪視規則之界定及標準化訪視之演練，以期提高訪員間的信度(interviewer reliability)。訪視工作係在各醫院的血液透析中心進行。由經過標準化的訪員就問卷內容對所有接受血液透析的末期腎衰竭患者進行標準化結構式詢問和解說。在訪視過程中，若受試者有睡覺、身體不適、情緒不穩、發生意外事件等干擾時，則停止訪問並擇期再訪。若連續二次均拒絕訪問者，則視為「拒訪者」，對於意識障礙(Glasgow昏迷指數在 15 分以下)、失明、身體狀況甚差(ECOG 等級為 4 級)的患者，則視為「無法受訪者」。

健康相關生活品質效用測量方法信度的評估

(一) 信度相關係數(Pearson's correlation coefficient)

信度即可靠性，係指測量結果的一致性

(consistency)或穩定性(stability)。健康狀態測量方法的信度通常可以再測信度(test-retest reliability)、觀察者內信度(intra-rater reliability)及觀察者間信度(interrater reliability)加以評估。

觀察者間信度(interrater reliability)加上再測信度(test-retest reliability)：由血液透析時間超過6個月的個案中，隨機選取15名病患，於訪視完成後間隔四週，由不同的訪員，以相同的訪視方式，對這些病患進行再測，以評估測量的信度。

觀察者內信度(intrarater reliability)加上再測信度：由血液透析時間超過6個月的個案中，隨機選取15名病患，於訪視完成後間隔6週，由相同的訪員，以相同的訪視方式，對這些病患進行再測，以評估測量的信度。

觀察者內信度加上再測信度加上不同的訪視方法的信度：由血液透析時間超過6個月的個案中，隨機選取15名病患，於訪視完成後間隔8週，由相同的訪員，以不同的訪視方式(電話訪談)，對這些病患進行再測，以評估測量的信度。

(二) 類內相關係數(intraclass correlation coefficient)

類內相關係數(38)可用於比較兩種不同測量尺度之測量方法的信度。其公式如下：類內相關係數 = 個體間的變異量 / (個體內的變異量 + 個體間的變異量 + 測量方法的變異量)。若一測量方法的類內相關係數較大，則該測量方法的測量信度較好。本研究中即分別比較 TTO 及 RS 兩種方法在不同訪視者相隔4週後以相同方法再測、相同訪視者相隔6週後以相同方法再測、相同訪視者相隔8週後以不同方法再測等三種情況下的類內相關係數，以評估其測量的信度。

健康相關生活品質效用測量方法效度的評估

效度係指一種測量工具或方法是否能真正測量到所想要測量的東西。效度愈高即表示測量的結果愈能顯現其所欲測量對象的真正特徵。今就 TTO 及 RS 二種測量方法之建構效度(construct validity)及效標相關效度(criterion-related validity)的評估方法分別說明如下：

(一) 建構效度(construct validity)：建構效度係指測量工具能測量理論的概念或特質的程度而言。在探討健康相關生活品質之測量方法的建構效度時，可利用輻合效度和辨別效度來考驗測量方法的建構效度[14,39]。本研究係利用不同的效用測量方法來測量健康相關生活品質，故僅能就輻合效度來探討測量方法的建構效度。

輻合效度(convergent validity)：輻合效度係指具有相同建構之不同測量方法所得的測量結果是相關的。本研究中分別以 SG、TTO、TTOVA 及 RS 等四種方法來測量病患的健康狀態效用數值，並探討各測量結果間的相關性，以評估其輻合效度。其中，TTO 及 RS 是所有的病患均進行測試，而 SG 及 TTOVA 則因人力及施測不易等限制，只分別隨機選取30名患者進行測試。

(二) 效標相關效度(criterion-related validity)[39]：效標相關效度係指某一測量工具與效標之相關或能測出效標之準確度。本

研究中將健康相關生活品質當成一個效用狀態來加以測量。SG 經Torrance研究證實在健康狀態之效用測量上具有很好的效度，並將其視為效標(criterion)來與其他效用測量方法所得的結果作比較[23]。本研究即以 SG 為效標，以評估 TTO 及 RS 等方法之效標相關效度。此外，本研究將病患依修改式卡式評分等級分組，並探討各分組的 TTO 及 RS 所測得的效用數值的分佈與修改式卡式評分等級間的相關性，以評估 TTO 及 RS 兩種測量方法是否具有同時效度(concurrent validity)。

資料分析

運用SAS套裝統計軟體6.03版[40]分析本研究之資料：

- (一) 以描述性統計值呈現人口統計學特性及臨床相關資料。
- (二) 以相關係數分析探討測量的效度。
- (三) 以相關係數分析，配對 t 檢定探討測量的信度。

結 果

回應率(response rate)

本研究中所有接受血液透析的末期腎衰竭患者共有148名，扣除9名無法訪問者(5名意識障礙者，1名失明者及3名身體狀況甚差無法回答者)，剩下的139人中，有14名拒訪，可被訪問的病人之完成訪問率為90%。拒訪者則由施測者觀察其活動情形，並以修改式卡式評分等級及意識狀態量表等兩個客觀性生活品質量表為依據，約略評估其健康相關生活品質。結果顯示14名拒訪者其修改式卡式評分等級及意識狀態量表的得分與完成訪問者間並無顯著差異。其拒訪的主要原因是先前已接受過類似研究的訪問或對本研究的問題不感興趣。

人口統計學特性及臨床相關資料

研究對象的人口統計學資料及臨床相關資料詳見林榮第等[41]所作之論文；本文僅略述如下：受訪者年齡平均為53.08歲。女性(53.6%)多於男性(46.4%)。教育程度以小學程

度者居多(30.4%)，次為高中或高職程度者(24.0%)。婚姻狀況以已婚、再婚居多(86.4%)。宗教信仰方面以佛教、道教居多(62.4%)。保險類別以勞農保居多(65.6%)。職業方面以未工作者(24%)最多，其次為從事農商工者(23.2%)、退休(22.4%)及家管(20%)。收入狀況以無收入居多(66.4%)。造成腎衰竭的原因以慢性腎絲球腎炎居多(55.2%)，其次為糖尿病(22.4%)。每週血液透析時數平均值(標準差)為10.94(2.19)小時。接受血液透析的年數平均值(標準差)為2.99(3.75)年。平均發病年齡為47.26±14.14歲。發病到住院間隔之平均值(標準差)為2.23(4.10)年，發病到血液透析間隔之平均值(標準差)為3.40(5.46)年。平均動脈壓的平均值(標準差)為96.91(13.51)mmHg。血比容平均值(標準差)為27.85(5.06%)，且大多數患者(76%)接受紅血球生成素注射。同時罹患數以同時罹患1種疾病居多(48%)，未同時罹患其他疾病者次之(36.8%)。24.8%患者同時患有心臟血管疾病，21.6%患者同時患有糖尿病。大多數患者(56%)不願意接受腎臟移植。

健康相關生活品質的效用數值

如表一所示，由 TTO、RS、SG，與 TTOVA 等效用測量方法所測得之血液透析患者自評健康相關生活品質的效用數值分別為0.61±0.26、0.57±0.16、0.75±0.24與0.61±0.23。此得分與完全健康者得分1.0相比較，顯示血液透析患者其健康相關生活品質有中等程度受到影響。

健康相關生活品質效用測量方法的信度

(一) 信度相關係數

觀察者間信度加上四週再測信度(表二)：TTO 及 RS (RS經log轉換)前後兩次測量值間的相關係數分別為0.93及0.91(0.97)，呈顯著的相關。TTO 及 RS 前後兩次測量值間經配對 t 檢定並無顯著差異。

觀察者內信度加上六週再測信度(表二)：TTO 及 RS (RS經log轉換)前後兩次測量值間的相關係數分別為0.87及0.79(0.77)，呈顯著的相關。TTO 及 RS 前後兩次測量值間經配對 t 檢定並無顯著差異。

觀察者內信度加上八週再測信度加上不同訪視方法的信度(表二)：TTO 前後兩次測量值間的相關係數為0.87，呈顯著的相關。RS 前後兩次測量值間並無顯著相關。TTO 及 RS 前後兩次測量結果間經配對 t 檢定並無顯著差異。

(二) 類內相關係數

不同訪視者相隔4週後以相同方法再測：TTO 及 RS 其類內相關係數分別為0.872及0.856。

相同訪視者相隔6週後以相同方法再測：TTO 及 RS 其類內相關係數分別為0.770及0.694。

相同訪視者相隔8週後以不同方法再測：TTO 及 RS 其類內相關係數分別為0.763及0.410。

健康相關生活品質效用測量方法的效度

(一) 建構效度

輻合效度：TTO 與 TTOVA、SG 及 RS

表一 健康相關生活品質的效用數值

測量方法	效用數值		個案數
	平均值±標準差	範圍	
時間交換	0.61±0.26	0.01~1.00	125
評價等級	0.57±0.16 (-0.61±0.36)	0.10~0.85 (-1.00~-0.07)	125
標準賭博	0.75±0.24	0.10~1.00	30
時間交換加上 視覺輔助工具	0.61±0.23	0.10~0.80	30

註：括號內數值表示評價等級數值經log轉換所得結果

表二 健康效用測量法之信度(reliability)

	觀察者間信度 加上 再測信度(4 週)		觀察者內信度 加上 再測信度(6 週)		觀察者內信度 加上 再測信度(8 週) 加上 不同訪視方法的信度	
	相關係數	前後差異 平均值±標準差	相關係數	前後差異 平均值±標準差	相關係數	前後差異 平均值±標準差
時間交換	0.93**	-0.021±0.11	0.87**	-0.022±0.108	0.87**	-0.018±0.122
評價等級	0.91** (0.97**)	-0.015±0.078 (-0.027±0.122)	0.79** (0.77**)	0.003±0.093 (0.007±0.176)	0.34 (0.28)	-0.050±0.166 (-0.09±0.309)
疲勞狀況分數	0.76**	2.067±3.127*	0.003	-0.733±5.637	-0.0218	2.267±6.681
身體症狀分數	0.72**	2.933±5.300*	0.59*	-0.400±5.068	0.29	1.200±7.692
疼痛狀態分數	0.75**	1.667±5.341	0.38	-0.133±8.114	0.55*	0.733±4.728
Barthel 氏評 分表分數	1.0**	0 ± 0	1.0*	0 ± 0	0.94**	0.33±1.29
修改式卡氏評 分等級	1.0**	0 ± 0	1.0*	0 ± 0	1.0**	0 ± 0
社會角色功能 分數	0.85**	0.067±0.594	0.79*	-0.733±0.884**	0.74**	0.933±0.884**
社會支持功能 分數	-0.12	0.200±0.676	0.14	-0.200±0.561	0.16	-0.133±0.640
人際接觸分數	0.53*	0 ± 2.268	0.56*	-0.267±2.690	0.30	1.267±2.685
精神健康狀態 分數	0.81**	-1.467±2.825	0.11	0.467±0.290	0.14	3.600±2.535*

註：個案數 = 15；前後差異之檢定採paired t test；括號內數值表示評價等級數值經log轉換所得結果

* 表示 P < 0.05

** 表示 P < 0.01

等其他三種效用測量方法所得結果間皆呈顯著的正相關(表三)。RS 與 TTO 及 TTOVA 等其他兩種效用測量方法所得結果間呈顯著的正相關。而 RS 所得效用數值經對數轉換後(見討論中說明)也與 SG 效用數值呈顯著的正相關(表三)。TTO 與 TTOVA、TTO 與 SG、TTO 與 RS，其測量所得結果的相關性呈依序遞減(表三)。RS 與 TTO、RS 與 SG，其測量所得結果的相關性也呈依序遞減(表三)。

(二) 效標相關效度：以 SG 當作效標，由研究發現，TTO 與 SG 測量所得結果間呈顯著相關(表三)。而 RS 效用數值在經對數轉換後與 SG 效用數值間也呈顯著相關(表三)。隨著修改式卡氏評分等級的增加，TTO 效用數值由零級的0.68逐漸遞減為三級的0.325，TTO 效用數值與修

改式卡氏評分等級間的相關係數為 -0.25，且達 $p < 0.01$ 的顯著水準(表四)。同樣的，隨著修改式卡氏評分等級的增加，RS 效用數值(RS經log轉換)由零級的0.66 (-0.43)逐漸遞減為三級的0.425 (-0.93)。RS 效用數值(RS經log轉換)與修改式卡氏評分等級間的相關係數為 -0.33 (-0.32)，且達 $p < 0.01$ 的顯著水準(表四)。

討 論

若一種測量方法可以反覆產生一致性的結果，則該測量方法是具有信度的。在健康狀態的測量上，通常以再測信度、觀察者內信度及觀察者間信度來評估測量方法的信度。一般而言，TTO 及 RS 皆具有不錯的觀察者內信度[23,33,42,43]、觀察者間信度[42]及六週(含)以內的再測信度[23,44,45]。

表三 效用(utility)測量法之輻合效度(convergent validity)

測量方法(1)	測量方法(2)	個案數	相關係數	P值
時間交換	時間交換加上視覺輔助	30	0.99	<0.0001
時間交換	標準賭博	30	0.86	<0.0001
時間交換	評價等級	125	0.58	<0.0001
		113*	0.66*	<0.0001*
標準賭博	評價等級	30	0.44	0.1024
			0.55**	0.0336**
評價等級	時間交換加上視覺輔助	30	0.79	<0.0001

註：* 表示除去時間交換數值自評為完全健康(1.00)，但臨床量表得分顯示有障礙的個案

**表示評價等級數值經log轉換

表四 用不同測量法對身體健康情況(以修改式卡式評分法作標準)之同時效度

效用測量法		修改式卡式評分等級			
		0	1	2	3
時間交換 (time trade-off)	個案數	23	84	14	4
	平均值	0.68	0.61	0.52	0.33
	標準差	0.12	0.26	0.26	0.28
	相關係數		- 0.25*		
評價等級 (rating scale)	個案數	23	84	14	4
	平均值	0.66	0.57	0.5	0.43
		(-0.43)	(-0.61)	(-0.81)	(-0.93)
	標準差	0.11	0.15	0.2	0.17
		(0.17)	(0.33)	(0.56)	(0.48)
相關係數		- 0.33* (- 0.32*)			

註：* 表示P < 0.01

括號內數值表示評價等級數值經log轉換所得結果

本研究因人力上的限制，僅分別於間隔4、6、8週由血液透析時間超過6個月的病程穩定個案中，隨機選取15名病患進行再測，若按ECOG等級分層統計分析，則將因各分層人數太少(再測患者大多集中於ECOG=1)，較無法顯示其原有的測量結果。故僅以表2的方式來表示信度的測量結果。除了RS在觀察者內信度加上八週再測信度加上不同訪視方法所得的信度稍差外，其餘的信度結果與其他學者研究的結果相近[23,33,42-45]。RS (RS經log轉換)在觀察者內信度加上八週再測信度

加上不同訪視方法所得的信度雖然稍差，但其前後兩次測量結果間的相關係數仍達0.34 (0.28)，只是未達 $p < 0.05$ 的顯著水準，且前後兩次測量結果經配對t檢定並無顯著差異。而在相隔8週後由相同訪員以電話訪談進行再測的信度之所以稍差，可能是由於「以電話訪談時缺乏視覺輔助工具來幫助病患作答」、「病人於電話訪談時，可能傾向於社交性作答」以及「病患對於健康狀態的喜好程度會隨時間的改變而改變」等因素所造成的。於間隔4、6、8週後實施再測，再測時間雖間隔較

長，但由於所施測對象均為意識狀態不受疾病影響且病程穩定之血液透析患者，且由多向度健康相關生活品質表的再測結果顯示(表二)[41,48]，在較穩定的Barthel氏評分表分數及修改式卡氏評分等級方面，病患的兩次測量結果幾乎是一致的。而在其他較具變動性的量表方面，病患的兩次測量結果間也有相當高程度的相關性。因此，病患在兩次測量期間內的健康相關生活品質是穩定的，並不會對 TTO及 RS 兩種健康狀態效用測量方法所得的結果造成影響。綜合上述結果我們發現，本研究中測量所得的結果具有很好的信度，在健康狀態的測量上，TTO及RS測量所得的結果較臨床量表穩定。對於不同的訪視者、不同的訪視時間、不同的訪視方法或訪視情境，TTO及RS所得的結果均很穩定，可見此兩種測量方法具有很好的信度。而根據類內相關係數的結果我們可以發現，TTO比RS具有較好的測量信度。

若一種測量方法可以精確地測量到所想要測量的東西，則該測量方法是具有效度的。效度一般分三種類型，即內容效度(content validity)、效標關聯效度(criterion related validity)與建構效度(construct validity)。其中以建構效度最為完善，部份測量專家認為建構效度包含其他二種效度。

本研究係利用不同效用測量方法來測量健康相關生活品質，故僅能就不同測量方法所得結果間的輻合程度(convergence)來探討測量方法的建構效度。輻合效度係指具有相同建構之不同測量方法所得的測量結果是相關的。以相同方法測量相同特質所得結果間的相關性應較不同方法測量相同特質所得結果間的相關性大。SG及TTO都是依據預期效用理論所發展出來的效用測量方法，而RS則是起源於心理計量學(psychometrics)的一種效用測量方法。SG是在具有風險的不確定情況下所進行的測量，而其餘效用方法則是在不具風險的確定情況下來進行測量。依理論及施測方法而言，TTO與TTOVA、SG及RS等方法間的相似性呈依序遞減；而RS與TTO及TTOVA較具相似性，而與SG較不具相似性。本研究顯示由SG、TTO、RS

與TTOVA等效用測量方法所得的結果間有顯著的相關性。這與Reed等人[46]之研究所得的結果一致。TTO所得結果與TTOVA、SG及RS等所得結果間的相關性隨著各組合間之測量原理及方法的相似性的遞減而減少。RS所得結果與TTOVA、TTO及SG等方法所得結果的相關性也隨著各組合間之測量原理及方法的相似性的遞減而減少。由上述結果可知，TTO及RS兩種方法均具有輻合效度，亦即此兩種效用測量方法在健康相關生活品質的測量上具有建構效度。

本研究中利用SG、TTO及RS所測得的血液透析患者之健康相關生活品質的效用數值與國外類似研究所得的結果相近[8,14]。Torrance(8)以SG為效標來評估TTO及RS的效度，結果發現TTO具有不錯的效度。因為RS是在不具風險的確定情況考量下所作的測量，故其結果須經轉換後，才具有不錯的效度[23,47]。轉換方式須視效用測量結果與效標間的函數圖形而定，本研究嘗試使用各種轉換方法(包括Torrance研究所用的冪方轉換)以期得到較好的結果。其中以經log轉換所得結果較佳，故採用此轉換方式。其所得結果與Torrance相似，唯RS的結果須經對數轉換後，才具有不錯的效標相關效度。

由本研究顯示，隨著修改式卡氏評分等級的增加(亦即健康相關生活品質逐漸變差)，TTO與RS兩種測量方法所得的效用數值皆呈現下降趨勢，且TTO及RS兩種方法所得的效用數值與修改式卡氏評分等級間具有顯著的相關性。亦即TTO與RS所得的效用數值對於修改式卡氏評分等級(健康相關生活品質)具有預測能力。故TTO及RS兩種方法在健康相關生活品質的測量上具有同時效度。Churchill[14]對末期腎病患者進行研究亦發現TTO效用測量方法對於接受不同透析類型之病患的健康相關生活品質的評估具有同時效度。

此外，由先前的研究[41,48]顯示，TTO及RS兩種測量方法所測得的健康相關生活品質效用數值與健康相關生活品質問卷三個向度所含各臨床量表之得分間可用線性複迴歸模式來顯示兩者間的關係，此亦表示TTO

及 RS 兩種方法在健康相關生活品質的效用測量上是具有效度的。

在本研究中將健康相關生活品質當成一個效用狀態來加以測量，雖然 SG 在健康狀態效用的測量上具有良好的信度及效度，但其施測方法太過於複雜，不易了解，對大多數受試者並不合適，而且施測的成本高，故不適合應用於大規模的研究。TTO 是 Torrance 等人[24]依據預期效用理論所發展出來，用於測量健康狀態效用的方法。與 SG 比較起來，TTO 具有良好的效度且較容易施測。其所得結果與 SG 相似，故適合用於測量健康狀態效用的研究上。但是當時間及效用間的函數不呈線性關係時，TTO 在使用上會受限制，而且 TTO 所測得的效用並不實際代表受試者是否接受治療的決定。RS 是起源於心理計量學(psychometrics)的一種測量方法，是最常用於測量健康狀態效用的方法。與 SG 比較起來，其所得結果若經轉換(通常是冪次方轉換；本研究中經對數轉換)，則具有不錯的效度。由於 RS 施測容易且成本低，故很適合應用於大規模的研究。

綜合上述結果，在健康狀態效用的測量方法中，SG、TTO 及 RS 皆具有良好的信度及效度，但就可行性(feasibility)而言，TTO 及 RS 是較受大多數學者所建議採行的方法。另外，由於 RS 施測容易，且所得結果與其他方法同樣具有信度及效度，故很適合應用於大規模的研究[8,14]。

誌 謝

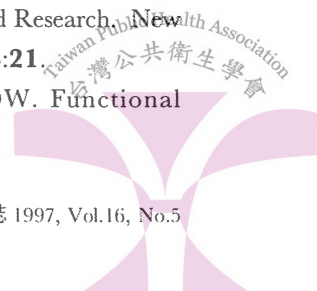
承蒙臺大醫院嚴潔鑫教授、蔡敦仁教授，新光醫院李素慧主任，和平醫院吳振龍院長、李壽星主任、范俊雄主任，臺大公衛所張珏教授、鄭泰安教授、曹昭懿老師、國防公衛所白璐教授，行政院衛生署及所有接受訪視的單位在本研究進行中給予協助，謹此誌謝。

參考文獻

1. Sonnenfeld ST, Walds DR, Lemieux JA,

- McKasick DR. Projections of national health expenditures through the year 2000. *Health Care Financ Rev* 1991; **13**:1-27.
2. Patrick DL, Erickson P. Health status and health policy: quality of life in health care evaluation and resource allocation. Oxford, England: Oxford University Press, 1993: **25**.
3. Dubos R. The state of health and quality of life. *West J Med* 1976; **125**(1):8-9.
4. Calman KC. Definition and dimensions of quality of life. In : Aaronson NK, Beckman J, eds. *The quality of life of cancer patients*. New York: Raven Press, 1987:1-9.
5. Hornquist JO. The concept of quality of life. *Scand J Soc Med* 1982; **10**:57-61.
6. Guyatt GH, Jaeschke R. Measurements in clinical trails: Choosing the appropriate approach. In: Spilker B eds. *Quality of Life Assessment in Clinical Trials*. New York: Raven Press, 1990: 37-46.
7. Wenger NK, Mattson ME, Furberg CD, Elinson J, eds. *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*. Washington DC: Le Hacq, 1984: 11-15.
8. Torrance GW. Utility approach to measuring health-related quality of life. *J Chron Dis* 1987; **40**(6):593-600.
9. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; **118**:622-629.
10. World Health Organization (WHO). The constitution of the World Health Organization. *WHO Chron* 1947; **1**:29.
11. Ware Jr JE. Standard for validating health measures: definition and content. *J Chron-Dis* 1987; **40**(6):473-480.
12. deHann R, Aaronson N, Limburg M, Langton Hewer R, van Cervel H. Measuring quality of life in stroke. *Stroke* 1993; **24**(2):

- 320-327.
13. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter O, Cox D. Quality of life measures in health care. I: Applications and issues in assessment. *Br Med J* 1992; **305**: 1074-7.
 14. Froberg DG, Kane RL. Methodology for measuring health-state preferences -II: Scaling methods. *J Clin Epidemiol* 1989; **42**(5): 459-71.
 15. Jennet B. Predicting outcome in individual patients after head injury. *Lancet* 1976; **1**:1081.
 16. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: Macleod CM eds. *Evaluation of chemotherapeutic agents*. New York: Columbia University press, 1949: 191-205.
 17. Zubrod GC, Scheidemann M, Frei E III, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide. *J Chron Dis* 1960; **11**:7-33.
 18. Patrick DL, Stein J, Porta M, Porter CQ, Ricketts TC. Poverty, health services and health status: lessons from rural America. *Milbank Q* 1988; **66**(1):105-36.
 19. Novello AC. Ethical, Social and Financial Aspects of End-Stage renal disease. In: Brenner BM eds. *The kidney*. 4th ed. Philadelphia: Saunders, 1991: 2424-2443.
 20. Ghavamian M, Gutch C, Kopp F et al. The sad truth about hemodialysis in diabetic nephropathy. *JAMA* 1982; **222**:1868-1889.
 21. Levin NW. Quality of life and hematocrit level. *Am J Kidney Dis* 1992; **20**(suppl1): 16-20.
 22. Von Neumann J, Morgenstern O. *Theory of games and economic behavior* Princeton, NJ: Princeton University Press, 1944.
 23. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences* 1976; **10**(3):128-36.
 24. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care program. *Health Service Res* 1972; **7**:118-133.
 25. Furlong W, Feeny D, Torrance GW, Barr R, Horsman J. *Guide to design and development of health-state utility instrumentation*. Hamilton, Ontario: McMaster University, CHEPA working paper series, 1990: 90-9.
 26. Torrance GW. Measurement of health-state utilities for economic appraisal. *A Rev J Health Economics* 1986; **5**:1-30.
 27. Campeau L. Letter to the Editor. *Circulation* 1976; **54**:522.
 28. Ferris BG. Epidemiology standardization project. *Am Rev Respir Dis* 1978; **118** (suppl):7-10.
 29. Laupacis A, Muirhead N, Keown p, Wong C. A disease-specific questionnaire for assessing quality of life in patients on hemodialysis. *Nephron* 1992; **60**:302-306.
 30. Brenner BM, Lazarus JM. Chronic renal failure. In: Brenner BM eds. *Harrison's principle of internal medicine*, 12th ed. New York, Mc-Graw Hill, 1991: 1150-1157.
 31. Shealy CN, Shealy M. Behavioral techniques in the control of pain: a case for health maintenance vs disease treatment. In: Weisenberg M, Tursky B, eds. *Pain: New perspectives in Therapy and Research*. New York: Plenum Press, 1976:21.
 32. Mahoney FI, Barthel DW. Functional



- Evaluation: the Barthel index. *Md state Med J* 1965; **14**(2):61-65.
33. Torrance GW. Preferences for health state: a review of measurement methods. In: Sinclair JC eds. *clinical and economic evaluations of perinatal programmes*, Mead Johnson Symposium on Perinatal and Developmental Medicine No. 2, Vail Colorado, 1982.
 34. Spitzer WO, Dobson AJ, Hall J et al. Measuring the quality of life of cancer patients. *J Chron Dis* 1981; **34**:585-597.
 35. Donald CA, Ware JE. The measurement of social support. In: Greenley JR eds. *Research in Community Mental Health*. Greenwich, Connecticut: JAI Press, Inc, 1984:325-370.
 36. Cheng TA. The design and development of a screening questionnaire (CHQ) for use in community studies of medical disorders in Taiwan. *Psychological Medicine* 1986; **16**: 415-422.
 37. Folstein MF. "Mini-mental state" a practical method for guiding the cognitive state of patients for the clinician. *J Psychiatr Rev* 1975; **12**:189.
 38. Snedecor GW, Cochran WG. *Statistical methods*, 6th ed. Iowa State University Press, Ames, 1971; 279-298.
 39. Patrick DL, Erickson P. Health status and Health Decisions. In: Patrick DL, Erickson P, eds. *Health status and health policy: quality of life in health care evaluation and resource allocation*. Oxford, England: Oxford University Press, 1993: 198-202.
 40. SAS Institution Inc. *SAS/STAT Users Guide*, release 6.03 edition. Cary, NC: SAS Institution Inc. ,1988.
 41. 林榮第、白璐、游芝亭、嚴潔鑫、王榮德：血液透析病人之健康相關生活品質問卷的設計與驗證。中華衛誌1996; **15**(4): 333-345.
 42. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. *Health services Res* 1973; **8**: 228-245.
 43. Cadman D, Goldsmith C. Construction of social value or utility-based health indices: The usefulness of factorial experimental design plans. *J Chron Dis* 1986; **39**:643-651.
 44. O'Connor AM, Boyd NF, Till JE. Influence of elicitation technique, position order and test-retest error on preferences for alternative cancer drug therapy. *Nursing Research : Science for Quality Care; Proc 10th National Nursing Research Conference*. Toronto: University of Toronto; 1985.
 45. Churchill DN, Torrance GW, Taylor DW et al. Measurement of quality of life in end-stage renal disease: the Time Trade-off approach. *Clin Invest Med* 1987; **10**:14-20.
 46. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preference for health outcomes: comparison of assessment methods. *Med Decis Making* 1984; **4**:315-329.
 47. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res* **30**:1043-1069.
 48. 林榮第：健康相關生活品質之效用測量方法的效度與信度的評估。國立台灣大學公共衛生學院碩士論文。1993;37p.

RELIABILITY AND VALIDITY OF UTILITY APPROACH TO MEASURING HEALTH RELATED QUALITY OF LIFE : AN EXAMPLE OF PATIENTS ON HEMODIALYSIS

RONG-DIH LIN¹ KAI-PING GRACE YAO²
CHIN-TIN YU³ JUNG-DER WANG³

The measure of health-related quality of life has been used progressively in clinical decision making, public health priority setting, assessing the effectiveness of health programs and economic cost-utility analysis. The purpose of this study is to evaluate the reliability and validity of two utility measurement techniques including time tradeoff and rating scale in the measure of health-related quality of life in order that we could develop an indicator for health policy decision and program evaluation based on Chinese cultural background. Subjects of this cross-sectional study consisted of all the patients of end-stage renal disease receiving maintenance hemodialysis at the hemodialysis room of three teaching hospitals in the Taipei between August 1993 and November 1993. Interviews were undertaken by standardized interviewers with a structured questionnaire. 125 subjects completed the interview. The response rate was about 90%. The results showed that both time tradeoff and

rating scale measurement techniques got consistent results by different interviewers and method and at different interviewing time and situations. These two measurement techniques had good performance in the assessment of validity including construct validity and criterion-related validity. The utility values measured by standard gamble, time tradeoff, and rating scale were 0.75, 0.61, and 0.57 respectively.

Standard gamble, time tradeoff and rating scale are all reliable and valid techniques in the measure of health-related quality of life. Based on feasibility, the time tradeoff and rating scale methods are more popular with most investigators. The rating scale method is easiest to administer and appears to yield results that are as valid as any other method. Thus, this would seem to be the method of choice in large-sample studies. (*Chin J Public Health. (Taipei): 1997; 16(5): 404-416*)

Key words: *healthy-related quality of life, utility, reliability, validity*

¹ Department of Internal Medicine, Taipei Municipal Ho-Ping Hospital, Taipei, Taiwan, ROC.

² Department of Psychology, National Taiwan University, Taipei, Taiwan, ROC.

³ Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University College of Public Health, Taipei, Taiwan, ROC.

